# PATIENT CLASSIFICATION USING ASSOCIATION MINING OF CLINICAL IMAGES

*Sumeet Dua[1,2], Vineet Jain[1] , and Hilary W. Thompson[2]*

[1]Data Mining Research Laboratory, Department of Computer Science, Louisiana Tech University, Ruston, LA USA. [2]Department of Ophthalmology, Louisiana State University Health Sciences Center, New Orleans, LA USA

## ABSTRACT

Automated clinical image data collection tools and apparatus are becoming increasingly important to the medical industry, and imaging databases are growing at an unprecedented rate. Consequently, grid-based telemedicine efforts require the autonomous classification of patient images from distributed sources for fast and accurate image storage, management, and retrieval. In this paper, we present a unique algorithm that performs feature discovery to find class-wise isomorphic Association Rules (ARs) among features. By discovering ARs, we are able to find unique and useful knowledge in images. To find knowledge, we first uniformly segment every image in a series and extract color and texture features for every segment., Next, we discover ARs for the color and texture features for image segments. We then exploit redundancy in the differentials of rule sets for the autonomous classification of patient image data with significant sensitivity and specificity. We demonstrate the efficacy of our approach with experimental results on a data set of diabetic retinopathy patients.

***Index Terms***— Classification, association, image databases, clinical decision support

## 1. INTRODUCTION

As digital instrumentation and computing resources continue to transform patient care, and, as data storage costs continue to fall, an abundance of larger, more complex patient databases are becoming available. Consequently, rapid deployment of isolated databases, often referring to the same patient, but stored in non-standardized data silos, i.e. disconnected desktops, enterprise servers, and departmental level storage systems, has recently become an issue of concern. The rise in telemedicine, translational medicine studies, and computational clinical decision support further escalate the need for (semi-) autonomous data integration and interoperability. Furthermore, the computerized medical record is on the verge of becoming practical in direct health-care applications, and will soon be a business necessity for health-care providers, further swelling data growth. This growth will lead to the renewal of demands for the development of novel technologies designed to organize and mine data and enhance biomedical and, consequently, computing research.

In this paper, we present a unique algorithm for the autonomous classification of ophthalmologic images using a data mining approach consisting of the four distinct components shown in Figure 1.

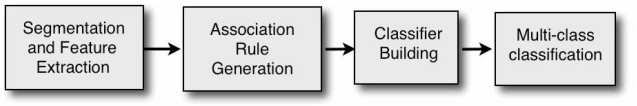



**Fig. 1.** The outline of the proposed classification schema.

A clinical image is autonomously classified when we assign signatures to it based on the ARs we discover for the color and texture features within the image and based on the comparisons we make between those features and the database of class-specific rule signatures. The sensitivity and specificity of this approach depends on how precisely the knowledge in the image is represented by the discovered signatures. Initially, the image is segmented using an equi-window based approach. For every segment, color and texture features are extracted, and ARs are discovered among the extracted features. These ARs are then organized for classification purposes.

## 2. SIGNIFICANCE AND BACKGROUND

Image classification can be applied in image annotation, disease and patient classification, content-based image retrieval, and information retrieval in large image databases. Mining novel feature content from images is a non-trivial task that involves the abstraction of various image details, desirably at several different levels of resolution, to obtain a unique image signature for data mining purposes. Several techniques for clinical and biomedical image classification have been proposed recently. These approaches could sort out images of respective disease types from a concentrated set of disease conditions. However, the approaches cannot be directly applied to autonomous patient classification because, for instance, patients could suffer from more than one condition or images could be captured at different stages of the same condition. Our work focuses on a current challenge to data mining in patient classification: novel techniques that can discover patient-sensitive patterns and exploit them for classification purposes with limited or no user supervision.

Medical image classification has received considerable attention in recent years. Ford *et. al.* [1] classified patients in three datasets, those with Alzheimer's disease, those with schizophrenia, and those with mild traumatic brain injury, against corresponding controls using fMRI activation maps, specifically, the statistical parametric mapping approach (contrast and t-maps). Hoi *et. al.* [2] successfully used "batch-

mode active learning" to classify multiple, non-overlapping images in five datasets (four UCI datasets and one medical image dataset). However, they did not use this approach to classify patients. Antonie *et. al.* [3] achieved a 70% accuracy rate in identifying abnormal breast cells by using AR mining in digital mammograms and were further able to break down abnormal cell images into fatty, glandular, and dense tissue masses to classify candidates for biopsy. Liow *et. al.* [4] achieved 69% and better results in the classification of HIV positive patients from healthy volunteers by comparing voxels and volume of interest based analyses in FDG PET brain scans.

On the other hand, AR discovery is a powerful data mining technique [5], but work in the application of ARs for medical image classification has been limited. Antonie *et. al.* [3] use ARs to classify MRI images of breast cancer. In their method, the image is divided into 16-equal parts, and four statistical features of pixel intensities: mean, variance, skewness, and kurtosis are extracted from each part. ARs are discovered using class labels as rules and are then employed for classification. In [6], the authors develop a new associative classification algorithm called Classification based on Atomic ARs (CAAR). CAAR only generates atomic rules under high, self-adaptive confidence thresholds and dynamic support thresholds. In CAAR, the image is segmented into *n x n* regions, and 19 features are extracted from these segments. The ARs used in this approach are limited to those that have a class-label as a consequent, and multiple passes are needed to classify the images in the testing stage. The proposed approach does not follow any such constraint.

## 3. METHODOLOGY

The proposed methodology consists of the four parts presented in Figure 1 on page 1: (1) Segmentation and Feature Extraction, (2) Data Preprocessing and Rule Generation, (3) Classifier Building, and (4) Multi-class classification. The first step, Segmentation and Feature Extraction, is performed as follows. The image data set is divided into test (40%) and training classes.(60%). We perform 5-fold cross validation in the testing phase to reduce any selection bias.

### 3.1. Image Segmentation and Feature Extraction

The image is segmented into overlapping square segments of size *n x n* windows. The user can supply *n* based on the degree of resolution expected in discovering coherent relationships between images of the same class. For each window in the image, RGB color components are converted to LUV feature set, which are then used as the three-color features ($C_1$, $C_2$, $C_3$). Gray level co-occurrence matrices are used to extract the texture of each segment in the image. The gray-level co-occurrence matrix is a two dimensional matrix of joint probabilities, $P_{d,r}(i,j)$ $i \in 0..n, j \in 0..n$, between pairs of pixels separated by distance $d$ in a given direction. The following statistical properties of these co-occurrence matrices [7] are used to derive texture features:

$$Energy(T_1) = \sum_i \sum_j P^2(i,j) \quad Entropy(T_2) = -\sum_i \sum_j P(i,j).\log P(i,j)$$

$$Contrast(T_3) = \sum_i \sum_j |i-j| P^2(i,j) \quad Homogeneity(T_4) = \sum_i \sum_j \frac{P(i,j)}{|i-j|}.$$

As a result of this step, each derived segment of training image $I_j$ is a tuple with 7 values ($C_1$,$C_2$, $C_3$, $T_1$, $T_2$, $T_3$, $T_4$)

### 3.2. Data Preprocessing

For effective AR discovery among the discovered features in the next step, the data needs to be preprocessed. Since the features of the segments are continuous, variable data types, quantitative ARs, rather than Boolean ARs are discovered [5].

```
Begin
// Read an image from a set of training images
// Remove those tuples from feature matrix where feature values
are NIL
   Not_clean ← find (feature, tuples with NIL values)
   feature ← delete Not_clean tuples
Arrange the image feature set in a matrix with each row
corresponding to a segment and the column corresponding to each
of the derived features for the segment.
// Discretize the continuous data by dividing the range into 10
intervals and substituting
// the values in one interval by the mean of the interval
FOR every feature (column) in feature matrix
  // find maximum and minimum for the feature
  max_f ←maximum(feature(f))
  min_f  ← minimum(feature(f))
  // use the maximum and minimum to divide the range into 10
  // intervals
  range ← max_f – min_f
  interval_f ← min_f with increment of (range/10) until max_f
  // Replace each value in that feature with the mean of the
  //interval it falls in e.g. (t+(t+1))/2 is substituted for any value
  // falling in interval t to t+1
  For each value in the feature
      Value_f ← mean(interval(t) to interval(t+1))
  EndFor
ENDFOR
//The result of the above step is a matrix of image features
// (Discretized_matrix)
```

**Fig. 2.** Algorithm for data preprocessing and discretization.

Value range is partitioned into eight *equi-width* intervals for each feature type. Each interval is mapped to a corresponding mean value to simplify AR discovery and exploit redundancy for classification purposes. The algorithm we use for this step is presented in Figure 2.

### 3.3. Association Rule Discovery

Association rule discovery is performed as follows [5]. Let $I$ denote the set of all images in a particular class. The features can then take a set of $k$ discrete values $\{v_1, v_2, ....., v_k\}$. We denote the value of feature $F_j$ for image $I_j$ by the symbol $I_j[F_j]$. For each image ($I_j$) and feature ($F_j$) and for each set $X$ of images, $X \subseteq I$, $p \subseteq \{1,2,....,k\}$ define the sets of $I_j$ $|F_i$ $I_j$ and $X$:

$present(I_j \mid F_i, p) := \phi$ if $I_j[F_i] \neq v_p$; {i} if $I_j[F_i] = v_p$,

$present(I_j, p) := \{i \mid I_j[i] = v_p\}$; and

$present(X, p) := \bigcup_{I \subseteq X} present(I, p)$.

We also define for some index set $P$ and some set of features, $\{F_i \mid i \subseteq P\}$, the present set of X given $\{F_i \mid i \subseteq P\}$ as follows: $present(X \mid \{F_i \mid i \subseteq I\}, p) := \bigcup_{i \subseteq P} \bigcup_{I_j \subseteq I} present(I_j \mid F_i, p)$. For $X \subseteq I$, $p \subseteq \{1,2,....,k\}$, we define p-support of X to be, $s(X, p) := \#\, present(X \mid \{F_i \mid i \subseteq P\}, p)$. For disjoint subsets X and Y of $I$, $p \subseteq \{1,2,....,k\}$, we write $X(p) \Rightarrow Y(p)$ to indicate that $X \cap Y = \text{ø}$ and $present(X, p) \subseteq present(Y, p)$. We refer to $X(p) \Rightarrow Y(p)$ as an Association Rule (AR). An AR has a *support*, $s(X(p) \Rightarrow Y(p))$, defined to be, $s(X(p) \Rightarrow Y(p)) := \{\, i \mid present(X \mid \{F_i \mid i \in P\}, p) \subseteq present(Y \mid \{F_i \mid i \in P\}, p) \,\}$. Finally, we define the *confidence* of $X(p) \Rightarrow Y(p)$ as follows: $c(X(p) \Rightarrow Y(p)) := [s(X(p) \Rightarrow Y(p))] / s(X, p)$. The rules are filtered by a minimum measure of support and confidence, which provide a parametric control on the redundancy and significance of these rules. The association rules provide a unique abstraction of features in an image by drawing relationships (associations) between them. These relationship features are derived and then exploited for classification as described in the next section, and their efficacy is evaluated.

### 3.4. Classification

The classification is performed as follows. Sixty percent of the data is used for training. The rules common to all the images for the same patient are extracted and called $R_{common}$'common rules.' During our experiments, this figure lies somewhere between 0 and 728, depending on how similar the images are. Feature extraction for each of the test images is performed, and $R_{test}$ rules are extracted for each test image. This rule set is compared with each training image of every class, and the number of matching rules between $R_{common}$ and $R_{test}$ per image compared to the test image. Let these rules be $R_x$ for each image comparison. Among the $R_x$ rules, we say $R_{x\text{-}common}$ belong to $R_{common}$ obtained above from the training set. Then,

*False dismissals ( FD%) = ( ($R_{common}$ - $R_{x\text{-}common}$)/$R_{common}$)*100*

*False alarm average = ($R_x$ - $R_{x\text{-}common}$)*

*False alarms (FA%) = (($R_x$ - $R_{x\text{-}common}$)/$R_x$)*100*

## 4. EXPERIMENTAL RESULTS

We performed three different sets of experiments to find the intra-class similarity of nine different patients (nine patient classes) suffering from either Non-proliferative Diabetic Retinopathy (NPDR) or Proliferative Diabetic Retinopathy (PDR). Figure 3 shows a representative example of such fundus images from one patient. For the experiment, ten images in each class were taken, and 60% were used for training. The remaining images were used for testing (querying). A total of 24 (6 X 4) comparisons were made for each image set. After rule extraction, false alarm and false dismissal rates were calculated as explained in the previous section. The results are presented in Table 1 below. To evaluate the contributions of most discriminatory combinations of features and their efficacy in generating effective ARs for classification, we calculated the false alarms and dismissals for combined sets of 2 and 6 features.

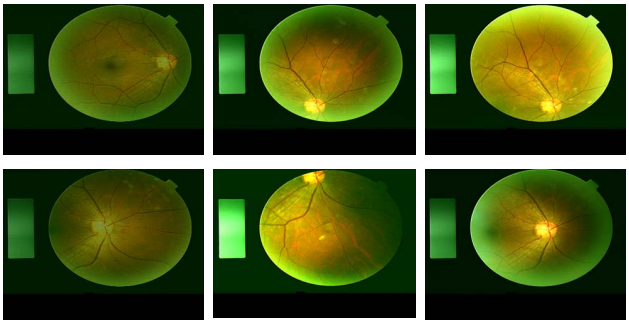


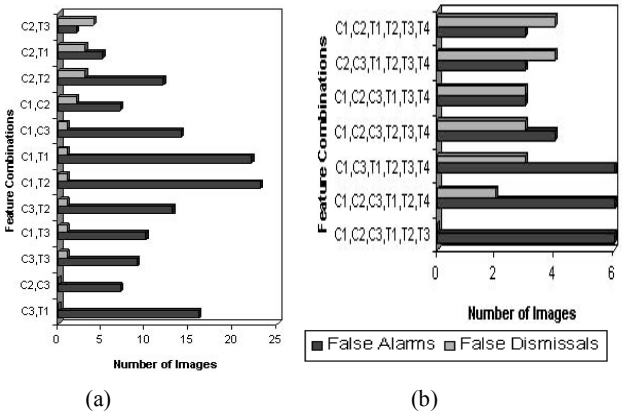
**Fig. 3.** Representative images of DR for one patient.

We observed that a combination of [*H*, entropy] and [*H*, *V*, energy, entropy] coefficients led to the highest rate of false alarms and that [*V*, energy] achieved the lowest rate of false dismissals. Our observations are presented in Figure 4.

**Table 1.** The classification results on a patient database.

| Patient set id | Common rules | FA (avg.) | FD (%) | FA (%) |
|---|---|---|---|---|
| *p01* | 42 | 455 | 0 | 30 |
| *p02* | 309 | 409 | 0.48 | 24 |
| *p03* | 4 | 420 | 0 | 33 |
| *p04* | 15 | 351 | 3.6 | 30 |
| *p05* | 15 | 465 | 0 | 36 |
| *p06* | 40 | 505 | 15 | 32 |
| *p07* | 728 | 114 | 0.14 | 9 |
| *p08* | 27 | 457 | 0.92 | 29 |
| *p09* | 671 | 101 | 0.4 | 8 |

Our approach can discover different levels of ARs. The 2-level rule has two ascendants (e.g. $C_2, T_1 \rightarrow T_4$), the 3-level rule has 3 ascendants (e.g. $C_1, T_1, T_4 \rightarrow T_3$), and so on.




(a) (b)

**Fig. 4.** Frequent feature combinations and matching images for (a) 2-feature, and (b) 6-feature combinations.

We measured the affect of different AR levels on classifier performance. As demonstrated in Figure 5, false dismissals rise rapidly when rule levels increase and false alarms decrease. We expect that an increase in rule levels will increase the specificity of the rule in discriminating between classes.

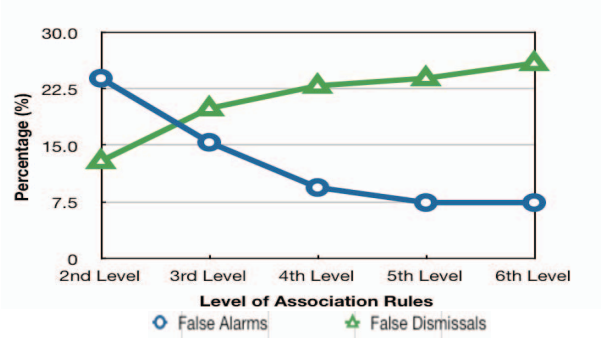



**Fig. 5.** Effect of levels of association rules on FD% and FA%

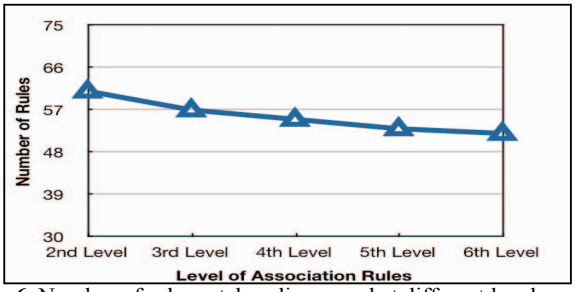



**Fig. 6.** Number of rule matches discovered at different levels.

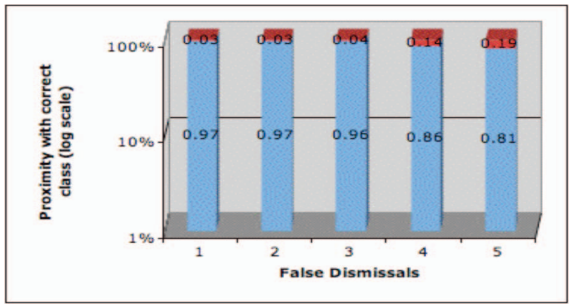



**Fig. 7.** Proximity of classified images with correct class.

Figure 6 shows the similarity (in terms of matching rules) we achieved by changing the level of rules. Figure 7 shows the proximity of false results with the correct class; a false dismissal generates a corresponding false alarm in another class. The labels in Figures 6 and 7 show how close the correct class is in the case of false alarms and false dismissals. The scores are computed as follows. If an image rejected as false dismissal has *x* similarity with the correct class and *y* similarity with another class (with closest match to the query), then the proximity with the correct class is *x/y*. The upper data labels present in Figure 7 show the difference *1 – x/y*.

## 5. CONCLUSIONS

Patient classification in medical imaging has a range of applications spanning both the biomedical and healthcare domains. The efficacies of such methods frequently rely on the discovery of effective and reliable feature extraction methods that can maximize the intra-class similarities. We have proposed a proof-of-principle schema for the discovery of ARs among features in images for autonomous classification of images to the patient of origin. Our extensive experimentation has shown that these ARs discover unique relationships among embedded features and possess discriminatory power. While the elucidation of scalability and portability of such isomorphic relationship discovery based frameworks is far from complete, we believe that this work will increase improved classification schemas for a variety of image domains and clinical data classification problems. Future work will include studying the variety of such image databases and developing scalable methods for feature discovery for superior classification.

## 6. REFERENCES


[1] J. Ford, H. Farid, F. Makedon, L.A. Flashman, T.W. McAllister, V. Megalooikonomou, and A.J. Saykin, "Patient Classification of fMRI Activation Maps," *Medical Image Computing and Computer-Assisted Intervention- MICCAI 2003*, Springer Berlin/Heidelberg, 58-65, 2003.

[2] S.C.H. Hoi, R. Jin, J. Zhu, and M.R. Lyu, "Batch Mode Active Learning and Its Application to Medical Image Classification," in *Proceedings of the 23rd International Conference on Machine Learning*, ACM, 2006, 417 – 424.

[3] M.-L. Antonie, O.R. Zaiane, and A. Coman, "Application of Data Mining Techniques for Medical Image Classification," in *Proceedings 2nd Int. Workshop Multimedia Data Mining,* MDM/KDD,, August 26, 2001.

[4] J.-S. Liow, K. Rehm, S.C. Strother, J.R. Anderson, N. Morch, L.K. Hansen, K.A. Schaper and D.A. Rottenberg, "Comparison of Voxel-and Volume-of-Interes-Based Analyses in FDG PET Scans of HIV Positive and Healthy Individuals," *Journal of Nuclear Medicine*, 612-621, 2000.

[5] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," in *Proceedings of the 20th Int'l Conference on Very Large Databases*, September 1994.

[6] X. Xu, G. Han, H. Min, "A Novel Algorithm for Associative Classification of Image Blocks," in *Proceedings of the 4th International Conference in Computer Information*, IEEE, 2004.

[7] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Trans. on SMC*, New York, 610–621, 1973.


## 7. ACKNOWLEDGEMENTS


The patient classification project was supported by Grant Number P20 RR16456 under the INBRE Program of the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH).